

The Mehrotra Predictor-Corrector Interior-Point Method
as a Perturbed Composite Newton Method

R.A. Tapia
Yin Zhang
M. Saltzman
A. Weiser

July, 1990
(revised July 1993)

TR90-17

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE JUL 1993		2. REPORT TYPE		3. DATES COVERED 00-00-1993 to 00-00-1993	
4. TITLE AND SUBTITLE The Mehrotra Predictor-Corrector Interior-Point Method as a Perturbed Composite Newton Method				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 21	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

The Mehrotra Predictor-Corrector Interior-Point Method as a Perturbed Composite Newton Method*

R. Tapia[†] Y. Zhang[‡] M. Saltzman[§] A. Weiser[¶]

July, 1990

(Revised July 1993)

*This research was supported in part by NSF Coop. Agr. No. CCR-8809615 and was initiated early in 1990 while the second and third authors were at Rice University as visiting members of the Center for Research in Parallel Computation.

[†]Department of Computational and Applied Mathematics and Center for Research on Parallel Computation, Rice University, Houston, Texas 77251-1892. This author was supported in part by AFOSR 89-0363; DOE DEFG05-86ER25017; and ARO 9DAAL03-90-G-0093.

[‡]Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21228. This author was supported in part by NSF Grant DMS-9102761 and DOE Grant DE-FG05-91ER25100.

[§]Department of Mathematical Sciences, Clemson University, Clemson, South Carolina 29634.

[¶]Center for Research on Parallel Computation, Rice University, Houston, Texas 77251-1892.

Abstract

It is well known that the celebrated Kojima-Mizuno-Yoshise primal-dual interior-point method for linear programming can be viewed as a damped perturbed Newton's method. Recently, Mehrotra suggested a predictor-corrector variant of this method. It is currently the interior-point method of choice for linear programming. The simplified Newton method, at the expense of fast convergence, reduces the work required by Newton's method by reusing the initial Jacobian matrix. The composite Newton method attempts to balance the trade-off between expense and fast convergence by composing one Newton step with one simplified Newton step. In this work we demonstrate that if the Newton component in the Kojima-Mizuno-Yoshise primal-dual method is replaced with a composite Newton component, then the resulting method is the Mehrotra predictor-corrector method.

1 Introduction

In subsection 1.1 we review the composite Newton method, in 1.2 we recall the primal-dual interior-point method, in 1.3 we present Mehrotra's predictor-corrector interior-point method, and in 1.4 we present our perturbed composite Newton interior-point method. Section 2 contains equivalence results between the Mehrotra predictor-corrector method and the perturbed level-1 composite Newton method. Since the level-1 composite Newton method is known to be cubically convergent, in Section 3 we study the cubic convergence aspect of the Mehrotra predictor-corrector interior-point method via our equivalence result. It is interesting to learn that the interior-point feature of the method, i.e., the step is damped so that iterates remain positive, precludes the standard proof of cubic convergence of the method. However, for nondegenerate problems it is possible to retain quadratic convergence. Recall that Zhang, Tapia and Dennis (1990) demonstrated that the primal-dual interior-point method can attain

quadratic convergence for nondegenerate problems. We then prove that by choosing steplength one in a neighborhood of the solution, cubic convergence can be attained by the predictor-corrector interior-point method for nondegenerate problems.

Numerical experimentation with the cubically convergent modification is most impressive and has been relegated to a companion paper, El-Bakry, Tapia and Zhang (1991), which numerically studies the local behavior of the predictor-corrector algorithm. Clearly an optimal implementation of the composite Newton interior-point method would allow m (the number of simplified Newton steps) to vary at each Newton step. This issue is not the subject of the current work, but probably merits further study. Finally, in Section 4 we give some concluding remarks.

1.1 The Composite Newton Method

Consider the nonlinear equation

$$F(x) = 0 \tag{1.1}$$

where $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$. By the *damped Newton method* for problem (1.1) we mean the iterative process

$$\begin{aligned} \text{solve } & F'(x_k)(\Delta x) = -F(x_k) \quad \text{for } \Delta x \\ \text{set } & x_{k+1} = x_k + \alpha_k \Delta x, \quad k = 0, 1, \dots \end{aligned} \tag{1.2}$$

The flexibility of being able to choose α_k less than one is important from global convergence considerations. When the choice of steplength is $\alpha_k = 1$ we drop the qualifier *damped*.

Under standard assumptions Newton's method is known to give Q -quadratic convergence. Not counting the work required to evaluate the function F or its Jacobian, the algebra required per iteration is $O(n^3)$, since the dominant task is the factorizing of the $n \times n$ Jacobian matrix $F'(x_k)$. For large n this can be a very serious concern.

A particularly obvious technique for reducing the amount of algebra needed at each iteration is given by the *damped simplified Newton method*

$$\begin{aligned} &\text{solve } F'(x_0)(\Delta x) = -F(x_k) \quad \text{for } \Delta x \\ &\text{set } x_{k+1} = x_k + \alpha_k \Delta x, \quad k = 0, 1, \dots \end{aligned} \tag{1.3}$$

The simplified Newton method requires an initial factorization of $F'(x_0)$ and then a solve at each iteration; hence it requires only $O(n^2)$ algebra per iteration. However, it gives only Q -linear convergence and it is not at all clear in what cases it should be preferred to Newton's method, since the slow convergence might force a prohibitive number of iterations.

In an effort to cover the middle ground between the extremes of Newton and simplified Newton it is very natural to consider the variant of Newton's method which takes m simplified Newton steps between every two Newton steps. By the *damped (level- m) composite Newton method* we mean the iterative procedure

$$\begin{aligned} &\text{solve } F'(x_k)(\Delta x_i) = -F(x_k + \Delta x_0 + \dots + \Delta x_{i-1}) \quad \text{for } \Delta x_i, i = 0, \dots, m \\ &\text{set } x_{k+1} = x_k + \alpha_k(\Delta x_0 + \Delta x_1 + \dots + \Delta x_m), \quad k = 0, 1, \dots \end{aligned} \tag{1.4}$$

Of course it is possible to introduce a different steplength control $\alpha_{k,i}$ for each correction Δx_i , $i = 0, \dots, m$; however we have no need to consider such flexibility.

It is reasonably well known that, under the standard Newton's method assumptions, the level- m composite Newton method has a Q -convergence rate of $m + 2$. A proof can be found in Chapter 10 of Ortega and Rheinboldt (1970). The damped level-1 composite Newton method where one Newton step is composed with one simplified Newton step is of particular interest to us. It can be written

$$\begin{aligned} &\text{solve } F'(x_k)(\Delta x_N) = -F(x_k) \quad \text{for } \Delta x_N \\ &\text{solve } F'(x_k)(\Delta x_S) = -F(x_k + \Delta x_N) \quad \text{for } \Delta x_S \\ &\text{set } x_{k+1} = x_k + \alpha_k(\Delta x_N + \Delta x_S), \quad k = 0, 1, \dots \end{aligned} \tag{1.5}$$

Ortega and Rheinboldt (1970) credit the cubic convergence of the level-1 composite Newton method to Traub (1964). However, the notion of composing Newton steps with simplified Newton steps is much older and a part of the folklore of Newton's method. It is generally felt by practitioners that the formulation of composite Newton steps is of value when n is large and the function F can be evaluated cheaply; this is clearly the situation for the primal-dual interior-point method for linear programming described in Subsection 1.2.

Observe that each level- m composite Newton iterate can be viewed as a major iterate and is the result of $m + 1$ inner iterations. The average amount of algebra per inner iteration is $O((n^3 + mn^2)/(m + 1))$ and is $O(n^2)$ for large m . The average convergence rate for the inner iterates is the $(m + 1)$ -st root of $m + 2$ and behaves like 1 for large m . It is no surprise then that for large m the level- m composite Newton method behaves like the simplified Newton method. It follows that an optimal implementation of composite Newton would not only vary m at each Newton step but would keep m relatively small.

1.2 The Primal-Dual Interior-Point Method

Consider a linear program in the standard form

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && Ax = b \\ & && x \geq 0 \end{aligned} \tag{1.6}$$

where $c, x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ ($m < n$) and A has full rank m .

The first-order optimality conditions for the linear program (1.6) can be written

$$F(x, y, \lambda) \equiv \begin{pmatrix} Ax - b \\ A^T \lambda + y - c \\ XYe \end{pmatrix} = 0, \quad (x, y) \geq 0 \tag{1.7}$$

where $y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^m$ are dual variables, $X = \text{diag}(x)$, $Y = \text{diag}(y)$, and $e^T = (1, \dots, 1) \in \mathbb{R}^n$.

The point (x, y, λ) is said to be feasible for problem (1.7) if $Ax = b$, $A^T\lambda + y - c = 0$, and $(x, y) \geq 0$. A feasible point (x, y, λ) is strictly feasible if $(x, y) > 0$. We tacitly assume that strictly feasible points exist.

It is now well understood how the primal-dual interior-point method introduced by Kojima, Mizuno and Yoshise (1989) can be stated in the framework of a damped and perturbed Newton's method applied to problem (1.7). In presenting this algorithmic framework we will write $z = (x, y, \lambda)$, $\Delta z = (\Delta x, \Delta y, \Delta \lambda)$, $\Delta X = \text{diag}(\Delta x)$, and $\Delta Y = \text{diag}(\Delta y)$. We also let $\min(u)$ denote the smallest component of the vector u and \hat{e} denote the vector $(0, \dots, 0, 1, \dots, 1)^T$ where the number of zeros is $n + m$ and the number of ones is n .

Algorithm 1 (Primal-Dual Interior-Point Method)

Given $z_0 = (x_0, y_0, \lambda_0)$ with $(x_0, y_0) > 0$, for $k = 0, 1, \dots$, do

$$(1) \quad \text{Solve } F'(z_k)(\Delta z) = -F(z_k) \text{ for } \Delta z_N \quad (1.8)$$

$$(2) \quad \text{Choose } \mu_k > 0 \text{ and}$$

$$\text{solve } F'(z_k)(\Delta z) = \mu_k \hat{e} \text{ for } \Delta z_c$$

$$(3) \quad \text{Set } \Delta z = \Delta z_N + \Delta z_c$$

$$(4) \quad \text{Choose } \tau_k \in (0, 1) \text{ and set } \alpha_k = \min(1, \tau_k \hat{\alpha}_k) \text{ where}$$

$$\hat{\alpha}_k = \min \left(\frac{-1}{\min(X_k^{-1} \Delta x)}, \frac{-1}{\min(Y_k^{-1} \Delta y)} \right) \quad (1.9)$$

$$(5) \quad \text{Set } z_{k+1} = z_k + \alpha_k \Delta z.$$

Actually in most implementations the formula (1.9) for α_k is further broken down and one steplength is used to update the x -variable and another is used to update the y -variable and the λ -variable. While this distinction is of value in practice, it is not an issue in the present work and consequently will be ignored.

Recently, under mild assumptions, Zhang, Tapia and Dennis (1990) demonstrated that for nondegenerate and degenerate problems Q -superlinear convergence could be attained by Algorithm 1 by merely letting $\sigma_k \rightarrow 0$ and $\tau_k \rightarrow 1$, where σ_k is defined by $\mu_k = \sigma_k x_k^T y_k / n$. Moreover, for nondegenerate problems Q -quadratic convergence could be attained by letting $\sigma_k = O(x_k^T y_k)$ and $\tau_k = 1 + O(x_k^T y_k)$. Zhang and Tapia (1991) showed that these results held under weaker assumptions.

The Newton step Δz_N defined in Step (1) can very likely point toward the boundary of the positive orthant, necessitating a very small choice for the steplength α . The major role of the centering step Δz_c defined in Step (2) is to remedy this situation. Hence it seems quite reasonable that the choice for the centering parameter μ_k should also be a function of the Newton step Δz_N . This is particularly true in delicate applications. For example, recently there was considerable speculation as to whether an instance of Algorithm 1 could have both polynomial complexity and superlinear convergence. In their original paper Kojima, Mizuno and Yoshise (1989) presented choices for τ_k and μ_k leading to polynomial complexity. Zhang, Tapia and Dennis (1990) presented conditions on τ_k and μ_k that guaranteed superlinear convergence. Zhang and Tapia (1990) settled this concern by demonstrating the existence of choices for τ_k and μ_k that guaranteed both polynomial complexity and superlinear convergence. Ji, Potra, Tapia and Zhang (1991) extended this result to linear complementarity problems. In both these applications the choice of μ_k depended strongly on the Newton step Δz_N . Hence, the centering step Δz_c had to be calculated as described in Step (2) above. However, in less delicate applications where μ_k depends only on z_k , Steps (1) and (2) in Algorithm 1 can be combined and the combined step $\Delta z = \Delta z_N + \Delta z_c$ can be obtained as the solution of

$$F'(z_k)(\Delta z) = -F(z_k) + \mu_k \hat{e}_k. \quad (1.10)$$

In this way the backsolve required by Step (2) can be saved. This is the more common presentation of the Kojima-Mizuno-Yoshise algorithm and is fine for restricted

applications.

The sense in which the primal-dual interior-point method can be viewed as damped perturbed Newton should be clear. The qualifier damped speaks to the steplength $\alpha_k \leq 1$ in Step (5). The qualifier perturbed speaks to the fact that the step Δz consists of the Newton step Δz_N perturbed by the centering step Δz_c ; see Step (3).

1.3 The Predictor-Corrector Interior-Point Method

Mizuno, Todd, and Ye (1989) suggested and studied an algorithm which they labeled a predictor-corrector algorithm. In their algorithm the predictor step is a damped Newton step for problem (1.7), producing a new strictly feasible iterate. The subsequent corrector step is a centered Newton step. In this corrector step, the choice of μ , the centering parameter, is based on the predictor step. Both the predictor and the corrector steps require essentially the same amount of work, namely, the evaluation and factorization of the Jacobian matrix.

Mehrotra (1989) later presented the following variant of Algorithm 1, which he also referred to as a predictor-corrector method. A common feature in these two predictor-corrector approaches is that the value of the centering parameter in the corrector step depends on the predictor step. However, unlike Mizuno, Todd and Ye's corrector step, Mehrotra's corrector step does not evaluate a fresh Jacobian matrix. Instead, it reuses the Jacobian matrix used by the predictor step. Recall that $\hat{e} = (0, \dots, 0, 1, \dots, 1)^T$.

Algorithm 2 (Predictor-Corrector Interior-Point Method)

Given $z_0 = (x_0, y_0, \lambda_0)$ with $(x_0, y_0) > 0$, for $k = 0, 1, \dots$ do

$$(1) \quad \text{Solve } F'(z_k)(\Delta z) = -F(z_k) \quad \text{for } \Delta z_p$$

- (2) Solve $F'(z_k)(\Delta z) = - \begin{pmatrix} 0 \\ 0 \\ \Delta X_p \Delta y_p \end{pmatrix}$ for Δz_M
- (3) Choose $\mu_k > 0$ and
solve $F'(z_k)(\Delta z) = \mu_k \hat{e}$ for Δz_c
- (4) Set $\Delta z = \Delta z_p + \Delta z_M + \Delta z_c$
- (5) Choose $\tau_k \in (0, 1)$ and set $\alpha_k = \min(1, \tau_k \hat{\alpha}_k)$ where (1.11)

$$\hat{\alpha}_k = \min \left(\frac{-1}{\min X_k^{-1} \Delta x}, \frac{-1}{\min(Y_k^{-1} \Delta y)} \right)$$
- (6) Set $z_{k+1} = z_k + \alpha_k \Delta z$.

While in the present section we are not concerned with the specific choice of the initial iterate z_0 or the various algorithmic parameters, we emphasize that Mehrotra suggested choices that allowed him to obtain very impressive numerical results.

1.4 The Perturbed Composite Newton Interior-Point Method

In this subsection we present our perturbed composite Newton interior-point method for problem (1.7). Recall that $\hat{e} = (0, \dots, 0, 1, \dots, 1)^T$. Our idea is to replace the Newton component in the primal-dual interior-point algorithm with a composite Newton component.

Algorithm 3 (Level- m Perturbed Composite Newton Interior-Point Method)

Given $z_0 = (x_0, y_0, \lambda_0)$ with $(x_0, y_0) > 0$ for $k = 0, 1, \dots$, do

- (1) Solve $F'(z_k)(\Delta z) = -F(z_k)$ for Δz_0
- (2) For $i = 1, \dots, m$ do
Solve $F'(z_k)(\Delta z) = -F(z_k + \sum_{j=0}^{i-1} \Delta z_j)$ for Δz_i (1.12)
- (3) Choose $\mu_k > 0$ and solve $F'(z_k)(\Delta z) = \mu_k \hat{e}$ for Δz_c

- (4) Set $\Delta z = \sum_{i=0}^m \Delta z_i + \Delta z_c$
- (5) Choose $\tau_k \in (0, 1)$ and set $\alpha_k = \min(1, \tau_k \hat{\alpha}_k)$ where
$$\hat{\alpha}_k = \min\left(\frac{-1}{\min(X_k^{-1} \Delta x)}, \frac{-1}{\min(Y_k^{-1} \Delta y)}\right)$$
- (6) Set $z_{k+1} = z_k + \alpha_k \Delta z$.

2 Predictor-Corrector as Perturbed Composite Newton

We say that two algorithms are equivalent if given a current iterate they produce the same subsequent iterate for the same choice of common algorithmic parameters.

Theorem 2.1 *The predictor-corrector interior-point method (Algorithm 2) is equivalent to the level-1 perturbed composite Newton interior-point method (Algorithm 3).*

Proof. Let $z = (x, y, \lambda)$ be the current iterate and let $\Delta z_p = (\Delta x_p, \Delta y_p, \Delta \lambda_p)$ be the predictor step for problem (1.7), i.e., Δz_p is obtained from Step (1) of Algorithm 2. By comparing Algorithm 2 with Algorithm 3 ($m = 1$), we see that our proof will be complete once we show that

$$F(z + \Delta z_p) = \begin{pmatrix} 0 \\ 0 \\ \Delta X_p \Delta y_p \end{pmatrix}. \quad (2.1)$$

Writing (2.1) in further detail gives

$$A(x + \Delta x_p) - b = 0 \quad (2.2)$$

$$A^T(\lambda + \Delta \lambda_p) + (y + \Delta y_p) - c = 0 \quad (2.3)$$

$$[x + \Delta x_p]_i [y + \Delta y_p]_i = [\Delta x_p]_i [\Delta y_p]_i, \quad i = 1, \dots, n. \quad (2.4)$$

By expanding we see that (2.4) is equivalent to

$$[y]_i[\Delta x_p]_i + [x]_i[\Delta y_p]_i = -[x]_i[y]_i, \quad i = 1, \dots, n. \quad (2.5)$$

However, (2.2), (2.3) and (2.5) are exactly the defining relations for the Newton step. Moreover, from the definition of Δz_p in Step (2) of Algorithm 2 it is clear that Δz_p is the Newton step. Hence (2.1) holds and we have established the equivalence. \square

While in Mehrotra (1989) no explanation for the predictor-corrector method is given, in a more recent paper, Mehrotra (1990), Mehrotra offers an interpretation of a related, but somewhat different, algorithm. Following the lead of Monteiro, Adler, and Resende (1988) he constructs a standard homotopy in a parameter, say δ , between problem (1.7) and a problem which had the current iteration as its solution. The primal-dual trajectory path parametrized by δ gives the solution of problem (1.7) for $\delta = 0$ and the current iterate for $\delta = 1$. He then views the iterate obtained from the predictor-corrector method as a point on a quadratic path which approximates the primal-dual trajectory path.

The equivalence represented by Theorem 2.1 was conjectured while listening to Mehrotra discuss his predictor-corrector method at the Second Asilomar Workshop on Progress in Mathematical Programming, Monterey, California, February 4-7, 1990. After proving Theorem 2.1 and while preparing this paper we received the paper of Lustig, Marsten, and Shanno (1990). In this paper the authors describe a comprehensive implementation of the Mehrotra predictor-corrector method and present impressive numerical results.

Lustig, Marsten and Shanno (1990) motivate Mehrotra's predictor-corrector method in the following manner. Rather than applying Newton's method to (1.7) to generate correction terms to the current iterate, they consider $F(z + \Delta z) = \mu \hat{e}$ directly, yielding

$$A(x + \Delta x) = b \quad (2.6a)$$

$$A^T(\lambda + \Delta\lambda) - (y + \Delta y) = c \quad (2.6b)$$

$$[x + \Delta x]_i [y + \Delta y]_i = \mu, \quad i = 1, \dots, n. \quad (2.6c)$$

Simple algebra reduces (2.6) to the equivalent system

$$A\Delta x = b - Ax \quad (2.7a)$$

$$A^T\Delta\lambda - \Delta y = c - A^T\lambda + y \quad (2.7b)$$

$$[x]_i [\Delta y]_i + [y]_i [\Delta x]_i = \mu - [x]_i [y]_i - [\Delta x]_i [\Delta y]_i, \quad i = 1, \dots, n. \quad (2.7c)$$

Observe that (2.7) defines the step $(\Delta x, \Delta y, \Delta\lambda)$ implicitly, i.e., in a nonlinear manner. In order to determine a step approximately satisfying (2.7) it seems reasonable to first solve $F(z) = 0$ for the Newton (predictor) step $(\Delta x_p, \Delta y_p, \Delta\lambda_p)$ and then use Δx_p and Δy_p on the right-hand side of (2.7) to solve for an “improved” step from (2.7) with this modified right-hand side.

It should be clear that the presentation (2.7), with Δx and Δy replaced by Δx_p and Δy_p only on the right-hand side reflects the level-1 composite Newton method corresponding to an unperturbed version written in the form

$$F'(z_k)\Delta z = -[F(z_k) + F(z_k - F'(z_k)^{-1}F(z_k))]; \quad (2.8)$$

while Mehrotra’s original presentation reflects form (1.5).

Lustig, Marsten and Shanno (1990) attempt an explanation of the predictor-corrector notion in terms of trajectories parametrized by the parameter μ . Their explanation contains some ambiguity in that it is not clear to what trajectories they are referring. Moreover, any explanation based on issues derived from μ cannot give a complete picture, since the predictor-corrector notion still makes sense even when the problem formulation is free of μ , i.e. $\mu = 0$ in all cases. However, implicit in these authors’ comments is the understanding that the corrector step can be viewed as a perturbed simplified Newton step.

3 Cubic Convergence

Much of the following analysis follows directly from material in Dennis and Schnabel (1983) or Ortega and Rheinboldt (1970). As before we consider problem (1.7) and use the notation $z = (x, y, \lambda)$. Also, recall that $\hat{e} = (0, \dots, 0, 1, \dots, 1)^T$ where the number of zeros is $n + m$ and the number of ones is n . The pure Newton method can be written

$$N(z) = z - F'(z)^{-1}F(z) \quad (3.1)$$

and the predictor-corrector interior-point method can be written

$$\hat{N}(z) = z - \alpha F'(z)^{-1}[F(z) + F(N(z)) - \mu \hat{e}]. \quad (3.2)$$

Therefore

$$\begin{aligned} \hat{N}(z) - z_* &= z - z_* - F'(z)^{-1}[F(z) + F(N(z))] \\ &\quad + (1 - \alpha)F'(z)^{-1}[F(z) + F(N(z))] + \alpha\mu F'(z)^{-1}\hat{e} \\ &= F'(z)^{-1}[F'(z)(N(z)) - F(N(z)) - F'(z)(z_*)] \\ &\quad + (1 - \alpha)F'(z)^{-1}[F(z) + F(N(z))] + \alpha\mu F'(z)^{-1}\hat{e} \\ &= -F'(z)^{-1}\{[F(N(z)) - F(z_*) - F'(z_*)(N(z) - z_*)] \\ &\quad + [F'(z_*) - F'(z)](N(z) - z_*)\} \\ &\quad + (1 - \alpha)F'(z)^{-1}[F(z) + F(N(z))] + \alpha\mu F'(z)^{-1}\hat{e}. \end{aligned} \quad (3.3)$$

Now, locally, i.e. in a neighborhood of the solution z_* , we know from standard Newton's method analysis that

$$\|N(z) - z_*\| = O(\|z - z_*\|^2).$$

Hence, we can rewrite the four terms on the right-hand side in (3.3) and obtain

$$\|\hat{N}(z) - z_*\| = O(\|z - z_*\|^4) + O(\|z - z_*\|^3) + |1 - \alpha|O(\|z - z_*\|) + \mu O(1);$$

which simplifies to

$$\|\hat{N}(z) - z_\star\| = O(\|z - z_\star\|^3) + |1 - \alpha|O(\|z - z_\star\|) + \mu O(1) . \quad (3.4)$$

In deriving (3.4) we used the fact that $\|F(z)\| = O(\|z - z_\star\|)$ and $\|F(N(z))\| = O(\|z - z_\star\|^2)$.

The term $\mu O(1)$ can be made $O(\|z - z_\star\|^3)$ by the choice of μ . Everything now hinges on the term $|1 - \alpha|O(\|z - z_\star\|)$. We must therefore take a very close look at the quantity $1 - \alpha$. Clearly, for cubic convergence, we need $|1 - \alpha|$ to be $O(\|z - z_\star\|^2)$.

Assuming strict complementarity, z_\star is a nondegenerate vertex solution, and z_k is feasible. Zhang, Tapia and Dennis (1990) obtained the useful expression

$$1 - \alpha_k = \frac{1 - \tau_k + \sigma_k \theta_k}{1 - \sigma_k \theta_k} + O(x_k^T y_k) \quad (3.5)$$

for the Newton interior-point method. See (3.7) of Zhang, Tapia and Dennis (1990). In (3.5), τ_k and σ_k are as in Algorithm 1, $\theta_k \in (\frac{1}{n}, 1]$ and $O(x_k^T y_k)$ is not necessarily zero and is exactly first order. Observe that $O(x_k^T y_k) = O(\|z - z_\star\|)$, since for feasible z_k we have $x_k^T y_k = \|F(z_k)\|_1$.

For the present purpose of studying $1 - \alpha_k$, the predictor-corrector primal-dual interior point method and the primal-dual interior-point method are philosophically the same, i.e., both can be viewed as perturbed Newton. In the former case the perturbation to the right-hand side of the defining relation is $\mu \hat{e} - F(z - F'(z)^{-1} F(z))$, while in the latter case the perturbation is merely $\mu \hat{e}$. Observe that these two perturbation terms differ by a term which is order $O(\|z - z_\star\|^2)$ or equivalently $O((x_k^T y_k)^2)$. Hence (3.5) is also valid for the Newton predictor-corrector interior-point method. It can now be seen from (3.5) that independent of the choices for τ_k and σ_k , the term $|1 - \alpha_k|$ is at best $O(\|z - z_\star\|)$ and the Newton predictor-corrector interior-point method, even for nondegenerate problems, cannot be shown to be cubically convergent by the standard approach. However, by choosing $\alpha_k = 1$ near the solution and

$\mu_k = O((x_k^T y_k)^3)$ we see from (3.4) that it is possible to obtain cubic convergence. We formally state these observations as the following theorem.

Theorem 3.1 *Let $\{x_k, y_k, \lambda_k\}$ be produced by Mehrotra's predictor-corrector interior-point method with z_0 strictly feasible. Assume*

- (i) *strict complementarity,*
- (ii) *x_* is a nondegenerate vertex, and*
- (iii) *$\{(x_k, y_k, \lambda_k)\}$ converges to (x_*, y_*, λ_*) .*

If the choices of σ_k and τ_k satisfy

$$0 \leq \sigma_k \leq \min(\sigma, c_1(x_k^T y_k)) \quad (3.6)$$

and

$$0 < \tau_k \leq \min(\tau, 1 - c_2 x_k^T y_k) \quad (3.7)$$

where $\sigma \in [0, 1), \tau \in (0, 1)$ and $c_1, c_2 > 0$, then the convergence is Q -quadratic, i.e. there exist $\gamma_2 > 0$ such that for k large

$$\|(x_{k+1}, y_{k+1}, \lambda_{k+1}) - (x_*, y_*, \lambda_*)\| \leq \gamma_2 \|(x_k, y_k, \lambda_k) - (x_*, y_*, \lambda_*)\|^2.$$

On the other hand, if instead of (3.6) we have

$$0 \leq \sigma_k \leq \min(\sigma, c_1(x_k^T y_k)^2) \quad (3.8)$$

and instead of (3.7) we have that for large k

$$\alpha_k = 1, \quad (3.9)$$

then the convergence is Q -cubic, i.e. there exist $\gamma_3 > 0$ such that for k large

$$\|(x_{k+1}, y_{k+1}, \lambda_{k+1}) - (x_*, y_*, \lambda_*)\| \leq \gamma_3 \|(x_k, y_k, \lambda_k) - (x_*, y_*, \lambda_*)\|^3.$$

Proof. The proof follows from combining the discussion given above with the details given in Zhang, Tapia, and Dennis (1990) for the proof of Theorem 4.1. \square

4 Concluding Remarks

In this paper we have studied the Mehrotra predictor-corrector philosophy and demonstrated that it is equivalent to the level-1 perturbed composite Newton philosophy.

We were intrigued by the discovery that, while the level-1 composite Newton method is known to be cubically convergent, this standard convergence rate proof applied to the predictor-corrector interior-point method gives at best quadratic convergence. The limitation of the standard proof results from the constrictive steplength choice forced on the method by the interior point philosophy, i.e., requiring the iterates to remain strictly feasible with respect to the nonnegativity constraints. We demonstrated that if one drops the interior-point aspect of the predictor-corrector method locally, i.e., in a neighborhood of the solution steplength one is selected, and also chooses the centering parameter to be of the order of the duality gap cubed, then cubic convergence can be attained for nondegenerate problems.

The research presented in Zhang, Tapia, and Dennis (1990), in Zhang, Tapia and Potra (1990), and the present research leads us to conjecture that we should implement Newton interior-point methods and their variants in a manner which near the solution sets the centering parameter to zero and takes steplength one, i.e., as old-fashioned Newton. Our preliminary numerical experiments employing this idea were impressive and motivated the more general study described in the companion paper El-Bakry, Tapia and Zhang (1991). The reader is referred to that paper for numerical results.

Acknowledgement. The authors thank an anonymous referee for critical comments that forced them to give a more consistent treatment of various conceptual notions.

References

- [1] J.E. DENNIS Jr. and R.B. SCHNABEL. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*.
Prentice-Hall, Englewood Cliffs, NJ, 1983. Russian edition, Mir Publishing Office, Moscow, 1988, O. Burdakov, translator.
- [2] A. EL-BAKRY, R.A. TAPIA and Y. ZHANG. Numerical comparisons of local convergence strategies for interior-point methods in linear programming. Technical Report TR91-18, Department of Mathematical Sciences, Rice University, Houston, Texas 77251-1892, September 1991.
- [3] M. KOJIMA, S. MIZUNO, and A. YOSHISE. A primal-dual interior-point method for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming, Interior Point and Related Methods*, Springer-Verlag, New York, 1989, pp. 29–47.
- [4] I.J. LUSTIG, R.E. MARSTEN, and D.F. SHANNO. On implementing Mehrotra’s predictor-corrector interior point method for linear programming. Technical Report SOR 90-03, Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey 08544, April 1990.
- [5] N. MEGIDDO. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming, Interior Point and Related Methods*, Springer-Verlag, New York, 1989, pp. 131–158.
- [6] S. MEHROTRA. On finding a vertex solution using interior-point methods. Technical Report 89-22, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, 1989. Revised January 1990.

- [7] S. MEHROTRA. On the implementation of a (primal-dual) interior point method. Technical Report 90-03, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, March 1990.
- [8] S. MIZUNO, M.J. TODD and Y. YE. Anticipated behavior of the path following algorithms for linear programming. Technical Report No. 878, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853, December 1989.
- [9] R.D.C. MONTEIRO, I. ADLER, and M.G.C. RESENDE. A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. ESRC Report 88-8, Department of Industrial Engineering and Operations Research, University of California, Berkeley, California 94720, March 1988.
- [10] J.M. ORTEGA and W.C. RHEINBOLDT. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, NY, 1970.
- [11] J. JI, F. POTRA, R.A. TAPIA and Y. ZHANG. An interior-point method with polynomial complexity and superlinear convergence for linear complementarity problems. TR91-23, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251-1892, July 1991. Submitted for publication.
- [12] J. TRAUB. *Iterative Methods for the Solution of Equations*. Prentice Hall, Englewood Cliffs, New Jersey, 1964.
- [13] Y. ZHANG and R.A. TAPIA. A superlinearly convergent polynomial primal-dual interior-point algorithm for linear programming. TR90-40, Department

- of Computational and Applied Mathematics, Rice University, Houston, Texas 77251-1892, June 1990. Also in *SIAM J. Optimization*, 3(1993), pp. 118-133.
- [14] Y. ZHANG and R.A. TAPIA. Superlinear and quadratic convergence of primal-dual interior-point algorithms for linear programming revisited. TR91-27, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251-1892, September 1991. Also in *JOTA*, 73(1992), pp. 229-242.
- [15] Y. ZHANG, R.A. TAPIA, and J.E. DENNIS. On the superlinear and quadratic convergence of primal-dual interior-point linear programming algorithms. Technical Report TR90-06, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251-1892, January 1990. Also in *SIAM J. Optimization*, 2(1992), pp. 304-324.
- [16] Y. ZHANG, R.A. TAPIA, and F. POTRA. On the superlinear convergence of interior-point algorithms for a general class of problems. Technical Report TR90-09, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251-1892, March 1990. Also in *SIAM J. Optimization*, 3(1993), pp. 413-422.